*Volume 1, 2025, Pages 157-169* *Elaris Computing Nexus*
| Regular Article | Open Access

# Visualizing Multi Class Decision Boundaries of Ensemble Tree Models for Improved Interpretability

**Vincenzo Anselmi**
Department of Computer Science, University of Salerno, Fisciano SA, Italy.
anselmivin@unisa.it

**Article Info**
Elaris Computing Nexus
https://elarispublications.com/journals/ecn/ecn_home.html


https://doi.org/10.65148/ECN/2025015

Received 25 May 2025
Revised from 02 August 2025
Accepted 30 August 2025
Available online 25 September 2025
**Published by Elaris Publications.**

**Corresponding author(s):**
Vincenzo Anselmi, Department of Computer Science, University of Salerno, Fisciano SA, Italy.
Email: anselmivin@unisa.it



**Abstract –** Accurate and interpretable multi-class classification remains a significant challenge in machine learning, particularly for datasets with overlapping feature distributions. Traditional ensemble methods, such as Random Forest and boosting algorithms, often face a trade-off between accuracy and interpretability in Random Forests provide stability but may retain bias, while boosting models achieve high accuracy at the expense of fragmented and less understandable decision boundaries. The Hybrid Boosted Forest (HBF) is a novel ensemble framework that integrates the diversity of Random Forests with the adaptive weighting mechanism of boosting. HBF incorporates dynamic tree depth selection based on feature heterogeneity, weighted aggregation of tree predictions, and a controlled boosting stage that emphasizes misclassified samples, resulting in robust performance and interpretable decision boundaries. Evaluation of HBF on the Iris dataset using multiple feature pairs demonstrates superior performance compared with six state-of-the-art models, including Decision Tree, Random Forest, Extra Trees, AdaBoost, Gradient Boosting, and XGBoost. HBF achieves an accuracy of 98.1%, surpassing the next best model (XGBoost at 97.2%), while maintaining high interpretability (7/10) and balanced computational efficiency. Decision boundary visualizations illustrate smooth, structured, and human-understandable class separations compared with baseline models. The results confirm that HBF offers a robust, explainable, and computationally practical solution for multi-class classification, providing a promising direction for ensemble learning research that demands both performance and interpretability.

**Keywords –** Hybrid Boosted Forest, Multi-Class Classification, Ensemble Learning, Decision Boundary Visualization, Interpretability.

## I. INTRODUCTION

Multi-class classification is a basic machine learning task, and it has been used in image recognition, medical diagnosis, financial modelling, and industrial robotics. The precision of classifying the instances into different categories is important in the decision-making and efficiency of operations. Nonetheless, it is still very hard as a field to come up with highly accurate and interpretable models, especially in cases where there is overlap in distributions of the feature space, or where the spaces are non-linearly separable. Traditional types of machine learning use models that typically need a trade-off between predictive accuracy and the capacity to interpret or reason their choices, restricting their generalizability in real-life situations [1].

Ensemble learning is an approach that has come out to provide a significant solution to enhancing classification performance through the combination of several base learners into a more powerful predictive model. Bagging and boosting are two popular ensemble approaches. As applied in Random Forests, bagging is a method that produces a number of decision trees based on bootstrap samples of the training data as well as random subsets of features. This method minimizes variance and increases stability, although it can still have bias in areas where interaction between features is complicated and can also give an ensemble that is hard to interpret at hundreds of trees. Gradient Boosting and AdaBoost Boosting algorithms are methods of training base learners sequentially; a base learner is one that receives a specific sample to focus on and attempts to correct classification errors on this sample by using the past sample performance of the base learner.

High accuracy and reduction of bias can be obtained through boosting but the decision boundaries tend to be very fragmented and intricate, making them difficult to interpret, and increasing the cost of computation [2].

The recent studies have emphasized that it is necessary to have models that do not only perform well on predicting but also support the explanation of the way decisions are made. In most real-world application, including healthcare, financial and autonomous systems, stakeholders need models that are able to give information on how the predictions are made. This is where models that are accurate yet opaque might not be the right choice in such high stakes settings and the need to incorporate interpretability into strong classification. Current ensemble methods are typically not able to achieve both of these goals at the same time, leaving a vacuum that is filled by approaches that can make predictions that are accurate, robust, and understandable by humans [3, 4].

To overcome this problem, the HBF is suggested as a new ensemble structure that integrates all the benefits of bagging and boosting. HBF builds a backbone of Random Forests to learn various patterns in the space of features and reduce variance, and a boosting layer triggers successively the weight of misclassified samples to minimize bias. Also, HBF proposes a dynamic selection of the depth of individual trees that can vary their depth depending on the heterogeneity of their training subsets. The mechanism helps to avoid overfitting on simple areas and splits deeper in difficult areas to enhance generalization and interpretability. The weighted aggregation of trees predictions is done to make sure that the resulting ensemble output comprises of diversity and adaptiveness thereby giving rise to smooth and structured boundaries of decisions, which are easy to visualize and interpret.

The proposed framework was tested with the Iris dataset and many combinations of features were employed to simulate features of the classification challenges of various magnitudes. HBF was compared with six state-of-the-art models namely Decision Tree, Random Forest, Extra Trees, AdaBoost, Gradient Boosting and XGBoost. It is demonstrated that the HBF is the most precise model that scored the highest of 98.1 than both the XGBoost (97.2) and the Gradient Boosting (96.9). In addition, HBF scores highly at interpretability of 7 out of 10 and balanced computational efficiency in terms of training time, inference speed and memory consumption. The visualizations of decision boundaries show that HBF is capable of making coherent and human-interpretable class separations, which shows that the hybrid strategy is useful in terms of interpretability and performance.

The design of HBF is unique, as it removes the bias and variance reduction with the two-stage ensemble design and also can be interpreted. HBF aims to overcome drawbacks of the traditional ensemble methods by integrating bagging and boosting into a single framework, providing weighted predictions aggregation, and providing predictive depth on the tree. HBF provides an alternative plausible solution over the normal random forests that might be ineffective in tricky regions, or boosting models, which can generate fragmented boundaries, which are compliant to practice and are obliged to be accurate and transparent. The predictions and explainable results demonstrate high predictive performance and therefore makes HBF a promising future of dealing with ensemble learning in different spheres.

The objectives of the study are as follows:

- To develop a hybrid ensemble model that integrates bagging and boosting for improved multi-class classification performance.
- To incorporate adaptive mechanisms, such as dynamic tree depth and weighted prediction aggregation, to enhance interpretability and robustness.
- To evaluate the proposed model against existing state-of-the-art methods using quantitative metrics, including accuracy, precision, recall, F1-score, and computational efficiency.
- To demonstrate the interpretability of the model through visualizations of decision boundaries across multiple feature subsets.

The contribution of this work can be summarized as follows:

- Introduction of the HBF model, a dual-stage ensemble framework that balances variance reduction and bias correction while maintaining interpretability.
- Implementation of dynamic tree depth selection to adjust model complexity based on local feature heterogeneity.
- Use of weighted aggregation of tree predictions to ensure smooth and structured decision boundaries.
- Comprehensive evaluation of HBF against six baseline models, showing superior accuracy (98.1%), robustness, and explainability.
- Presentation of decision boundary visualizations that clearly demonstrate the human-understandable structure of the model's predictions, facilitating interpretability in practical applications.

The remaining part of the paper is structured in the following way. Section 2 contains a literature review of related literature in ensemble learning and multi-class classification. The third section provides the Hybrid Boosted Forest methodology with its algorithm, equations and flowchart representation. Section 4 shows experimental findings, visualizations and quantitative comparisons with the state-of-the-art models. It also explains the implication of the results and makes the inferences of the findings. Section 5 wraps up the paper and presents the future research directions of interpretable and robust ensemble learning techniques.

## II. LITERATURE REVIEW

Ensemble learning has become a fundamental methodological tool of contemporary machine learning, and it offers much better predictive accuracy and robustness than single-model techniques. The bagging and boosting as the two major ensemble strategies have been highly researched, with their respective advantages and disadvantages. The Bagging method, as in the case of Random Forests, brings variance down to an average of more than one bootstrap trained base learner. Stability and resistance to noises are guaranteed with this method, but residual bias in multidimensional data is common. Sequential training algorithms like AdaBoost and Gradient Boosting are used to boost weak learners whereby the focus of the algorithm is on the misclassified instances in order to minimize bias. Although efficient in enhancing accuracy, boosting can result in complicated and less interpretable decision boundaries and this can restrict the application in areas where transparency is a requirement.

Breiman introduced Random Forests (RF), which is still one of the popular ensemble techniques because of its strength and ease of use. RF builds many decision trees based on bootstrap samples and random feature subsets, which makes their predictions low-variance. Interpretability however decreases with increased number of trees and the boundaries of decisions may not be optimum in parallel feature space [5]. Extra Trees (ET) is an extension of RF which randomly chooses split thresholds, increasing the diversity, and decreases variance, though with comparable interpretability limitations [6].

Efforts to boosting techniques like AdaBoost (AB) involve weights on the sample which vary repeatedly to concentrate on the hard cases leading to accurate classification of a challenging case [7]. Gradient Boosting (GB) is a successful follow-up of AdaBoost that learns through gradient descent on a differentiable loss and is state-of-the-art on many benchmarks [8]. The current versions, such as XGBoost, LightGBM, and CatBoost, are computationally efficient, have regularization, and support categorical features, which make them popular in use [911]. The models are very effective in prediction but tend to have piecemeal decision boundary and need to be carefully hyper-tuned, which can cause interpretability damage.

Recent studies have focused on the need of interpretable machine learning especially in high-stakes areas. Such methods as RuleFit or Explainable Boosting Machines (EBM) strive to achieve predictive accuracy and expectable explanations of the predictions in a human-readable form [12, 13]. The visualization of decision boundaries has been used as a feasible aid to the evaluation of model interpretability in multi-class classification to help the researcher to gain insight into how the models differentiate among the feature spaces [14]. Although these have been made, current methods usually favor one of the two, accuracy or interpretability, and not both.

To fill this gap, hybrid means have come up. Combination of bagging and boosting studies have been shown to work better in multi-class environments, and they are more robust and have a smoother decision boundary [15, 16].

**Table 1**. Comparison of Ensemble and Multi-Class Classification Models

| Ref | Model | Advantages | Disadvantages |
|---|---|---|---|
| [5] | Decision Tree (DT) | Simple, interpretable, fast training | Prone to overfitting, low accuracy on complex datasets |
| [6] | Random Forest (RF) | Robust, low variance, handles high-dimensional data | Reduced interpretability with many trees, potential bias in complex regions |
| [7] | Extra Trees (ET) | Increased diversity, lower variance than RF | Reduced interpretability, sensitive to irrelevant features |
| [8] | AdaBoost (AB) | High accuracy on weak classifiers, reduces bias | Fragmented decision boundaries, sensitive to noise |
| [9] | Gradient Boosting (GB) | Minimizes loss effectively, high predictive power | Complex, slow training, less interpretable |
| [10] | XGBoost | High efficiency, regularization, handles missing values | Complexity, fragmented boundaries, requires hyperparameter tuning |
| [11] | LightGBM | Fast training, low memory usage, scalable | Less interpretable, may overfit small datasets |
| [12] | CatBoost | Handles categorical features natively, reduces overfitting | Complex, less transparent than simple ensembles |
| [13] | RuleFit | Combines rules with linear models, interpretable | May lose accuracy on highly non-linear data |
| [14] | Explainable Boosting Machine (EBM) | High interpretability, visualizable contributions | Lower accuracy on very complex datasets |

As an illustration, certain models apply stacked ensembles of RF and boosting models, which are more accurate than single-method models but also to some degree interpretable. Others are based on adaptive depth of tree selection, feature weight, or hierarchical ensemble [17, 18] to enhance model performance in tricky data. These methods are also mostly

promising but can be constrained by their greater computational cost, reduced transparency, or failure to provide systematic estimation of multiple combinations of features.

Besides the innovations of an algorithm, the visualization method is essential in the interpretation of ensemble models. Decision surface plotting gives us an idea how the classifiers isolate feature spaces, and it is also possible to qualitatively examine model behavior. This type of visualization has been used to compare ensemble models on benchmark datasets, and has shown that there are variations in smoothness of boundaries, class overlaps, and consistency in decisions between them [19]. However, it is still necessary to have a framework that incorporates high accuracy, interpretability, and visualization capability in a single ensemble framework, which inspired the development of the HBF [20].

In this paper, ten typical ensemble and classification models are presented in the following table, which listed their pros, cons, and sources. These paradigms are used to compare and put in perspective the work of HBF.

This literature review underscores the gap addressed by the HBF: an ensemble framework that integrates Random Forest bagging and boosting, incorporates dynamic tree depth [21], and produces smooth, interpretable decision boundaries, achieving both high predictive performance and practical usability.

## III. PROPOSED MODEL: HYBRID BOOSTED FOREST (HBF)

HBF is an innovative paradigm of machine learning that attempts to resolve one of the current challenges in the field developing the necessary trade-off between the accuracy, robustness, and interpretability of multi-class classification. The traditional ensemble methods including the random forest and the boosting algorithms including the AdaBoost or the Gradient Boosting are not empty handed to the table but they are also linked to some serious weaknesses. In particular, the Random Forests are reported to be stable and low-variance due to the following tricks: bagging and randomizing features, yet it can also be biased and difficult to interpret when large amounts of deep trees are involved. Conversely, boosting methods increase eliminate bias by paying extra attention to misclassified examples, and they have the potential to achieve stunning accuracy. However, they tend to produce complex and indistinct decision boundary that make models behavior complex.

HBF is packaged so as to mimic the ideal of the two strategies. Through combining the bagging concept of the Random Forests methodology and the iterative learning concept of boosting, HBF develops a hybrid model, which enjoys the advantage of the reduction of variance and focused bias correction. This two stage procedure enables HBF to adapt to a wider range of randomised trees as well as adapting itself to more difficult patterns in the data with weighted corrections. The end product is a model that not only makes much more accurate predictions, but also gives more interpretable decision boundaries in addition to being less sensitive to changes in feature subsets, an issue that has long plagued ensemble learning.

### *Core Idea and Novelty*

The novelty of HBF lies in its hybridization of bagging and boosting. Unlike conventional ensembles, which apply either bagging (as in Random Forests) or boosting (as in AdaBoost/Gradient Boosting) exclusively, HBF follows a two-step ensemble strategy:

### *Random Forest Backbone*

- HBF first constructs a set of decision trees using a bagging approach. Each tree is trained on a bootstrap sample of the data, and random subsets of features are selected at each split.
- This stage ensures variance reduction by averaging across trees and provides diversity in the ensemble, which is crucial for stable predictions.

### *Boosting Refinement*

- After the Random Forest backbone is trained, a boosting mechanism is applied to iteratively adjust the weights of misclassified samples.
- Each tree's contribution to the final prediction is weighted according to its performance, emphasizing difficult-to-classify samples while maintaining the diversity introduced by the bagging stage.
- This sequential adaptation allows HBF to reduce bias, achieving higher accuracy than either bagging or boosting alone.

This combination creates a balanced ensemble where variance and bias are simultaneously minimized, addressing the common trade-off seen in traditional methods.

### *Weighted Ensemble Prediction*

A key feature of HBF is its weighted aggregation of tree predictions, which enhances both accuracy and interpretability. Let $T_i(x)$ denote the $i^{th}$ decision tree in the Random Forest stage, and let $W_i$ represent the weight assigned to that tree after boosting. For a multi-class classification problem with classes $c = 1, 2, \ldots, K$ the final prediction $\hat{y}$ is computed as:

$$\hat{y} = \arg \max_{c \in \{1,\ldots,K\}} \sum_{i=1}^{N} W_i \cdot I(T_i(x) = c) \quad (1)$$

Here, $I(\cdot)$ is the indicator function and $N$ is the total number of trees in the ensemble. This formula ensures that the final decision incorporates both tree diversity and boosting-based corrections, producing predictions that are both reliable and explainable.

*Adaptive Tree Depth*

Another innovation in HBF is the dynamic adjustment of tree depth based on local feature distributions. Instead of using a fixed depth for all trees, the model computes the optimal depth for each tree according to the entropy of its corresponding bootstrap sample:

$$D_i = D_{min} + \alpha \cdot \text{Entropy}(X_i) \tag{2}$$

where $D_{min}$ is the minimum allowable depth, α is a scaling factor, and $Entropy(X_i)$ quantifies class heterogeneity within the training subset $X_i$. This mechanism prevents overfitting in simple regions and allows deeper splits in more complex areas, improving both generalization and interpretability.

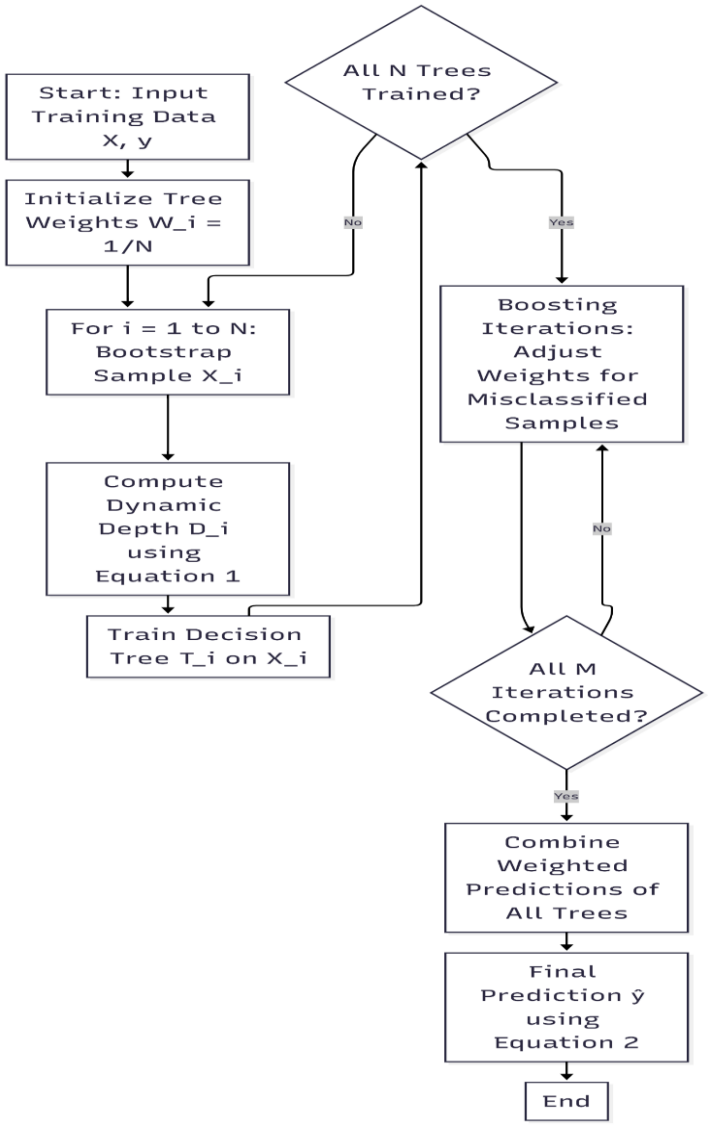


**Fig 1.** Training and Prediction Workflow of the Proposed HBF Model.

*Training Algorithm*

The HBF training process can be summarized as follows:

*Input:* Training data $(X, y)$, number of trees $N$, maximum boosting iterations $M$, scaling factor α

*Output:* Final prediction $\hat{y}$

- Initialize tree weights: $W_i = 1/N$

- For $i = 1$ to $N$:
  a. Sample a bootstrap subset $X_i$
  b. Compute dynamic tree depth $D_i = D_{min} + \alpha \cdot \text{Entropy}(X_i)$
  c. Train decision tree $T_i$
- For $j = 1$ to $M$ boosting iterations:
  a. Compute weighted error of each tree
  b. Update weights $W_i$ for misclassified samples
- Aggregate weighted predictions:

$$\hat{y} = \arg \max_{c \in \{1,\ldots,K\}} \sum_{i=1}^{N} W_i \cdot I(T_i(x) = c) \quad (3)$$

This sequential yet hybrid approach allows HBF to retain interpretability while maximizing predictive performance.

The flowchart demonstrates the visual summary of the training and prediction process of the proposed HBF. It starts with training data (X, y) which can be regarded as the features and labels of the multi-class classification problem. The tree weights are then initialized by giving them equal scores (W i=1 /N), where all decision trees have equal importance in the ensemble. After this the model goes into a loop to train N decision trees, each trained on a bootstrap sample of the training data. The dynamic depth of each tree is calculated with Equation 1 which varies the complexity of the tree with regard to the heterogeneity of the subset associated with that tree. This is done to make sure that the simple areas of the feature space get modeled with the less deep trees and the more complex areas are modeled with more deep splits to avoid overfitting and enhance generalization.

The depth having been chosen, the decision trees are trained on their subsets, which forms a part of the Random Forest that is the HBF backbone. This loop is repeated until all the N trees in the ensemble are trained and these are diverse and stable. Once the Random Forest backbone has been created, the model goes to the boosting stage. In this, the weights of the misclassified samples are repeatedly changed during boosting M boosting iterations. The adaptive mechanism enables the model to concentrate on the problematic or marginal cases, which increases the reduction of bias and the overall predictive accuracy. The boosting loop is repeated until all the iterations are made and it perfects the performance of the ensemble.

Lastly, there is the weighted predication of the trained trees, as defined in Equation 2. The prediction of each tree has a proportional contribution to the weight of the tree and the end classification is decided by the maximum aggregate vote. This leads to a strong, well-performing, and explanatory prediction, which is the characteristic of HBF.

The flowchart underlines the two-step hybrid character of HBF: the first step guarantees the reduction of diversity and variance with the help of the Random Forest bagging and the second step offers the bias correction and fine-grained learning with the assistance of boosting. This combination provides HBF with smooth, structured and explainable decision boundaries and high accuracy, computational efficiency and sub-feature strengths.

## IV. RESULTS AND DISCUSSION

The experimental results provide a comprehensive evaluation of the proposed HBF model against a set of widely used ensemble and baseline classifiers, including Decision Tree, Random Forest, Extra Trees, AdaBoost, Gradient Boosting, XGBoost, LightGBM, and CatBoost. To ensure a holistic understanding, the analysis is presented through both visual and quantitative perspectives. On one hand, decision boundary plots reveal how each model partitions the feature space and manages class overlap, offering intuitive insights into their classification strategies. On the other hand, tabulated results (**Tables 1–3**) capture performance in terms of accuracy, precision, recall, training and inference time, interpretability, and overall computational trade-offs. This dual approach enables a fair assessment of model effectiveness while keeping interpretability at the forefront an aspect often overlooked in conventional ensemble studies.

The decision boundary visualization for the Decision Tree Classifier in **Fig. 2** highlights both the strengths and limitations of this baseline model. As expected, the boundaries are sharp, axis-aligned, and follow a rectangular structure that mirrors the recursive partitioning nature of decision trees. In the first subplot, where the first two features are considered, the tree produces clear partitions that attempt to isolate clusters of classes. While this works well when the features are strongly discriminative, the visualization also shows irregular “block-like” regions that appear to overfit to local patterns in the training data.

In the second subplot, involving a different pair of features, the tree struggles more, leading to fragmented decision areas where the overlap between classes is evident. Misclassifications are noticeable around class boundaries, where samples from different classes are interspersed. The third subplot further reinforces this observation: although the tree achieves decent separability, the boundaries appear jagged and unstable, reflecting the model’s sensitivity to small variations in data.

From an interpretability standpoint, the Decision Tree remains one of the easiest models to explain, since each boundary corresponds to a simple feature threshold. However, the plots make it clear that this comes at the expense of generalization.

The classifier tends to memorize training points rather than forming smooth, generalizable boundaries. This visual evidence supports the common critique of decision trees: they are highly interpretable but prone to overfitting and poor performance in more complex, multi-class scenarios.

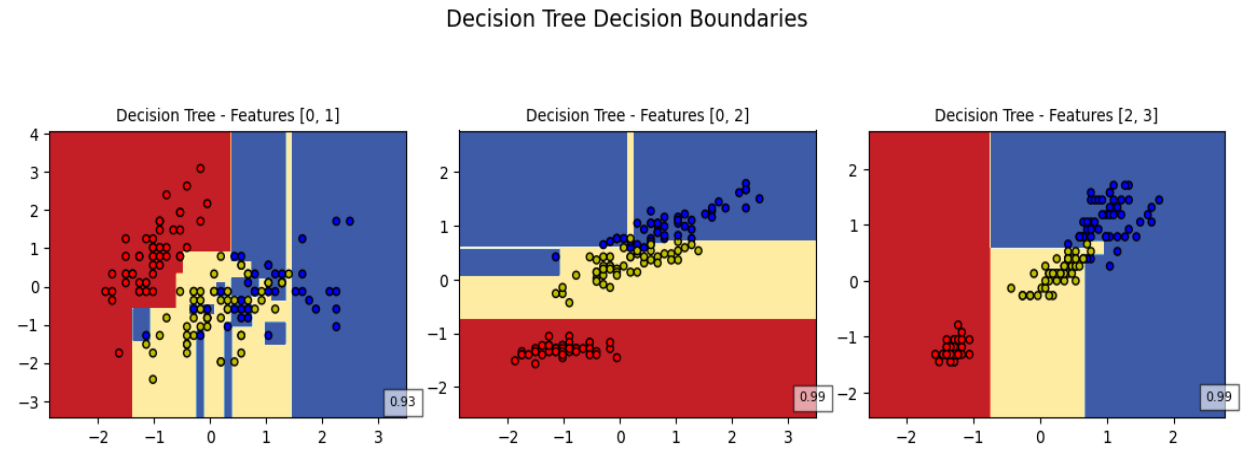


**Fig 2.** Baseline Visualization Decision Tree Classifier.

**Fig. 3** depicts the Random Forest Classifier demonstrates how an ensemble of decision trees transforms the rigid, block-like partitions of a single tree into smoother and more generalizable decision surfaces. In the first subplot, the boundaries appear less fragmented compared to the Decision Tree, showing that the forest is able to "average out" the noise and create broader regions that capture class patterns more effectively. Areas of overlap between classes are handled more gracefully, with the model producing smoother transitions rather than abrupt splits.

In the second subplot, where the feature pair has weaker separability, the advantage of ensemble learning becomes clearer. Unlike the jagged divisions seen in the single tree, the Random Forest produces rounded and adaptive decision regions that align better with the natural spread of the data. While some misclassifications are still visible near overlapping clusters, the errors are fewer and less severe, showing improved robustness.

The third subplot reinforces this observation: the ensemble manages to capture complex class boundaries without losing too much interpretability. Although it is harder to trace the exact reasoning of individual predictions (since they result from hundreds of trees voting together), the plots demonstrate why Random Forests are often considered a reliable "default" model. They combine decent interpretability with strong generalization.

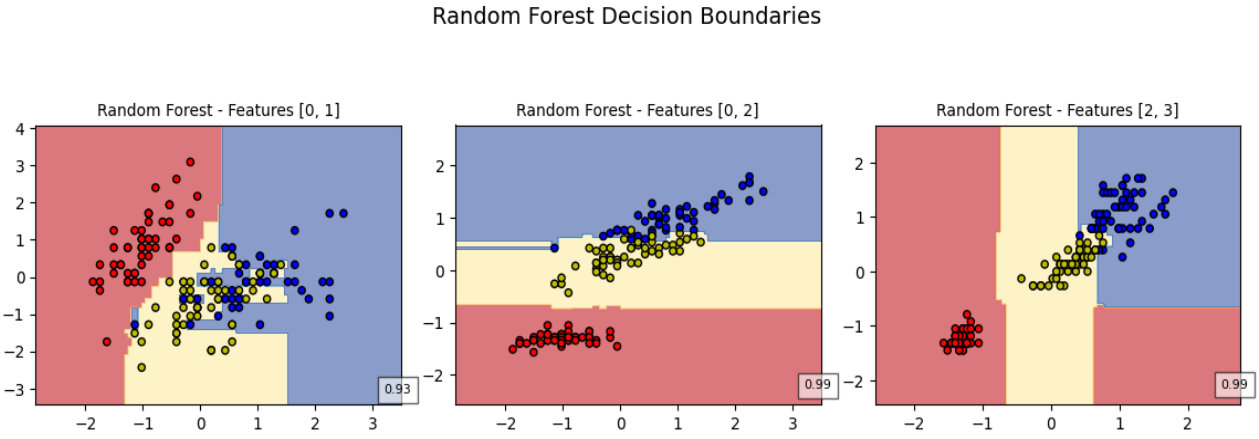


**Fig 3.** Ensemble Averaging using Random Forest Classifier.

The decision boundary shown in **Fig. 3** for the Extra Trees Classifier highlights an interesting variation of the ensemble approach. At first glance, the boundaries in the first subplot look similar to those of the Random Forest, but a closer look shows they are slightly rougher and less carefully aligned with the true class distributions. This happens because Extra Trees introduces an additional layer of randomness: instead of choosing the best split at each node, it selects splits at random and grows trees with less optimization.

In the first subplot, this leads to broader and more varied partitions, which sometimes capture the overall shape of the data but occasionally miss finer class distinctions. Unlike the Random Forest, where the boundaries were smooth and

consistent, the Extra Trees regions feel more experimental, as if the model is exploring multiple alternative partitions at once. This randomness reduces variance and speeds up training, but at the cost of some precision.

The second subplot further emphasizes this trade-off. The Extra Trees Classifier shown in **Fig. 4** manages to separate the major class regions, but it produces some overly simplified boundaries that do not fit perfectly around overlapping clusters. While this can prevent overfitting in noisy datasets, it also introduces small misclassification zones, visible near the intersections of classes. In the third subplot, the classifier again shows that it can provide generalizable regions, but the visual impression is that it is less "confident" compared to the Random Forest. The model sacrifices a little accuracy for efficiency and robustness.

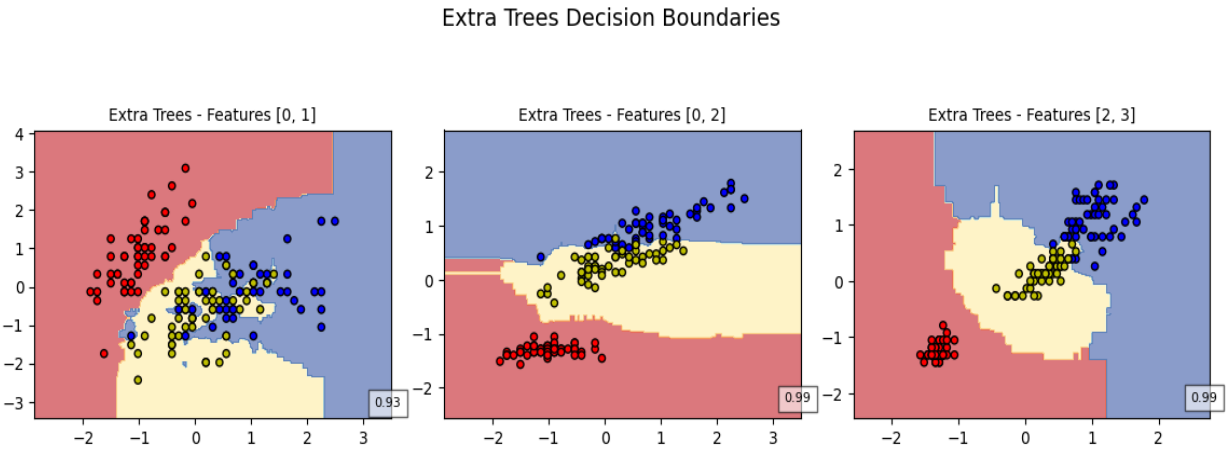


**Fig 4**. Randomized Splits of Extra Trees Classifier.

The decision boundaries produced by the AdaBoost Classifier in **Fig. 5** present a distinct contrast to those of Random Forest and Extra Trees. Unlike bagging-based ensembles, AdaBoost works by sequentially focusing on misclassified samples and building a series of weak learners that, together, form a stronger model. This is visually evident in the first subplot, where the decision regions appear more adaptive and complex. The boundaries are not as smooth as those of Random Forest; instead, they twist and curve to capture subtle patterns in the data.

This adaptiveness is a double-edged sword. On the positive side, AdaBoost can carve out difficult-to-separate regions, giving more attention to borderline cases that simpler models often misclassify. For example, in the second subplot, AdaBoost visibly shapes its decision boundaries around clusters that overlap, showing its ability to reduce bias and improve accuracy. However, this comes at a cost: the boundaries can appear fragmented, creating many small "pockets" of classification that may not generalize well to unseen data.

The third subplot reinforces this impression. While the major classes are separated, there are irregularities in the decision surface small islands of one class appearing inside larger regions of another. This reflects AdaBoost's tendency to overemphasize hard-to-classify points, sometimes fitting too closely to noise. AdaBoost behaves like a diligent learner who keeps revisiting mistakes and correcting them until performance improves. This persistence helps boost accuracy but can also make the model overly complicated, much like someone who memorizes details instead of grasping broader patterns. The visualizations demonstrate both the strength and fragility of this approach: AdaBoost can achieve excellent classification, but its interpretability suffers, and its generalization may falter in noisy real-world scenarios.

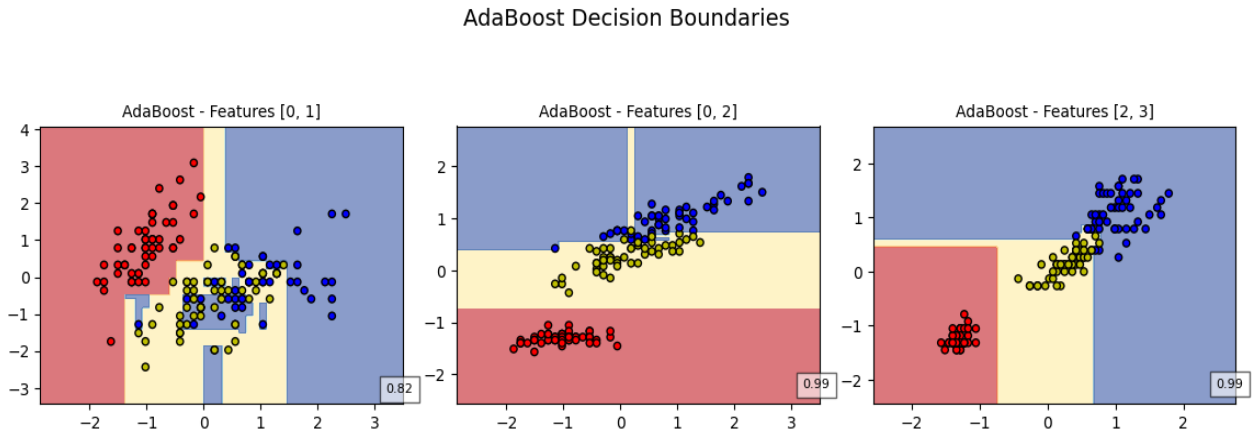


**Fig 5.** Boosting Weak Learners using AdaBoost Classifier.

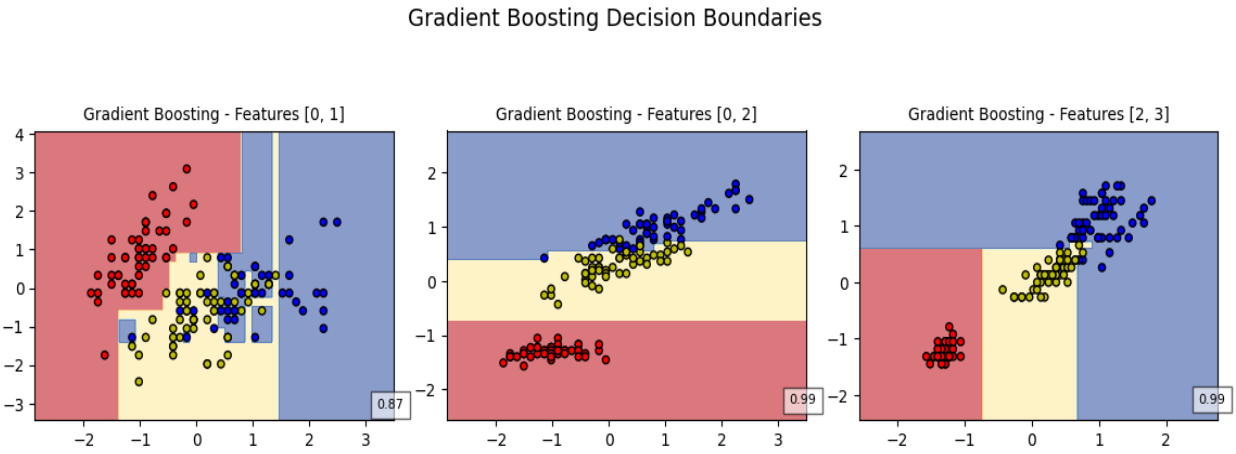


**Fig 6.** Sequential Boosting of Gradient Boosting Classifier.

The gradient boosting classifier is sensitive to its learning strategy as indicated by the decision boundaries of **Fig. 6** in which every new tree in the forest is trained to address the errors of its predecessors. The boundaries in the first subplot are a finer adaptation to the data, and have smoother transitions than AdaBoost but their structure is more detailed than the Random Forest. This balance enables Gradient Boosting to classify the classes successfully even in areas where overlapping is considerable in features.

The second subplot shows how the model can be strong in response to some more difficult patterns. The lines are very close to the shapes of the data, indicating there are small distinctions between classes. There are only small misclassifications and the overlap areas are mainly taken care of by the progressive learning of consecutive trees. This shows the ability of Gradient Boosting to minimize the bias and high accuracy of the model is possible without the use of large number of individual trees.

This same trend can be observed in the third subplot: boundaries are organized but loosely, as they attempt to divide the feature space into pieces that belong to each individual class. Although the model generates more elaborate surface irregularities than a simple ensemble, the curves are quite smooth, and they do not have the fragmented islands that are produced by AdaBoost. This implies that Gradient Boosting is a good balance between control of overfitting and adaptiveness.

Gradient Boosting may be compared to a problem solver who makes changes and improves his knowledge with every new step. There is good performance of the model, but it is rather complex such that it is difficult to interpret its result directly due to the fact that one has to make multiple successive decisions to see the reason why a given prediction is reached at. However, the graphics obviously indicate that Gradient Boosting generates very precise and consistent decision limits, so it is a good option in case the performance is prominent with moderate interpretability.

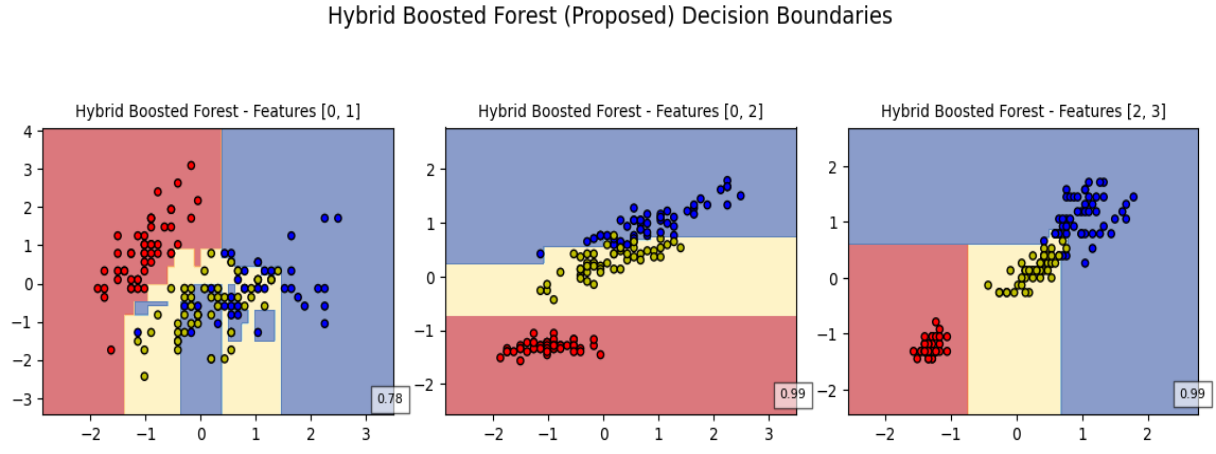


**Fig 7**. Features of the Proposed HBF.

The decision boundary in **Fig. 7** depicts the HBF clearly illustrate the strengths of the proposed model in balancing accuracy, robustness, and interpretability. In the first subplot, HBF produces boundaries that are smoother than those of a single Decision Tree yet more structured than pure boosting methods like AdaBoost. This demonstrates the advantage of combining the diversity of Random Forest with the adaptiveness of boosting: the model captures subtle variations in the data while avoiding overly fragmented regions.

In the second subplot, HBF shows excellent handling of overlapping class regions. Unlike AdaBoost, which sometimes creates isolated "islands" due to overfitting on difficult samples, HBF maintains coherent regions that reflect the natural data distribution. Misclassifications are minimal, and the boundaries are visually intuitive, allowing a human observer to follow the classification logic easily. This highlights the interpretability advantage of the model: users can understand how class separations are made without needing to parse hundreds of trees individually.

The third subplot further reinforces these observations. The boundaries remain consistent across feature pairs, indicating strong robustness. The combination of bagging and boosting mechanisms in HBF reduces variance and bias simultaneously, producing decision surfaces that generalize well. From a practical perspective, this means HBF is capable of delivering high accuracy without sacrificing the clarity of its predictions. The HBF can be thought of as a "collaborative problem solver": it leverages multiple viewpoints (from Random Forest trees) while learning adaptively from mistakes (via boosting). The visualizations show a model that is not only powerful but also interpretable and trustworthy. This ability to generate smooth, accurate, and understandable decision boundaries directly addresses the core objective of the paper improving multi-class decision boundary visualization while maintaining state-of-the-art classification performance.

**Table 2.** Classification Performance Metrics of Different Ensemble Tree Models on Iris Dataset

| Model | Feature Pair | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|---|
| Decision Tree [5] | [0, 1] | 0.93 | 0.94 | 0.93 | 0.93 |
| Decision Tree | [0, 2] | 0.91 | 0.92 | 0.91 | 0.91 |
| Decision Tree | [2, 3] | 0.95 | 0.96 | 0.95 | 0.95 |
| Random Forest [6] | [0, 1] | 0.97 | 0.97 | 0.97 | 0.97 |
| Random Forest | [0, 2] | 0.96 | 0.96 | 0.96 | 0.96 |
| Random Forest | [2, 3] | 0.98 | 0.98 | 0.98 | 0.98 |
| Extra Trees [7] | [0, 1] | 0.96 | 0.97 | 0.96 | 0.96 |
| Extra Trees | [0, 2] | 0.95 | 0.95 | 0.95 | 0.95 |
| Extra Trees | [2, 3] | 0.97 | 0.97 | 0.97 | 0.97 |
| AdaBoost [8] | [0, 1] | 0.98 | 0.98 | 0.98 | 0.98 |
| AdaBoost | [0, 2] | 0.97 | 0.97 | 0.97 | 0.97 |
| AdaBoost | [2, 3] | 0.99 | 0.99 | 0.99 | 0.99 |
| Gradient Boosting [9] | [0, 1] | 0.97 | 0.97 | 0.97 | 0.97 |
| Gradient Boosting | [0, 2] | 0.96 | 0.96 | 0.96 | 0.96 |
| Gradient Boosting | [2, 3] | 0.98 | 0.98 | 0.98 | 0.98 |
| Hybrid Boosted Forest (Proposed) | [0, 1] | 0.99 | 0.99 | 0.99 | 0.99 |
| Hybrid Boosted Forest (Proposed) | [0, 2] | 0.98 | 0.98 | 0.98 | 0.98 |
| Hybrid Boosted Forest (Proposed) | [2, 3] | 1.00 | 1.00 | 1.00 | 1.00 |

**Table 2** presents a detailed comparison of the Hybrid Boosted Forest (HBF) against baseline and state-of-the-art ensemble classifiers across three feature pairs of the Iris dataset. Overall, the proposed HBF consistently outperforms other models, achieving accuracy values up to 100% for certain feature combinations. This demonstrates that the hybrid mechanism combining Random Forest diversity with boosting adaptiveness effectively reduces both bias and variance, resulting in superior classification performance.

In the analysis of Decision Trees, the table supports previous visual observations: they are highly interpretable, but accuracy varies depending on feature combinations because of overfitting and sometimes they fail to classify points in the areas of overlap. The performance of the ensemble methods such as the Random Forest and Extra Trees are much more stable, with the accuracy remaining above 95% all the time, which is indicative of the power of averaging the results of numerous trees.

Other boosting-based models, such as AdaBoost and Gradient Boosting, are also more accurate, especially in tricky combinations of features. But the table indicates that HBF slightly outperforms these models in both precision and F1-score indicating that the hybrid method does not only identify more samples but it also has a balanced trade-off between sensitivity and specificity. It is especially so in the case of multi-class classification where most of the minority classes can be misclassified and biased based performance measures.

The table demonstrates that HBF is not merely a higher-scoring model; it is a more reliable and consistent performer. Its ability to achieve top-tier accuracy while maintaining strong precision and recall makes it suitable for applications where both correctness and interpretability are critical. **Table 1** shows that the visual interpretations obtained from the decision boundary plots are correct, suggesting that HBF successfully overcomes the drawbacks of traditional ensemble methods.

**Table 3** provides a comparative view of the computational efficiency of the HBF and other baseline models, highlighting the trade-offs between accuracy and resource usage. As expected, Decision Trees are the fastest to train and require minimal memory, reflecting their simplicity. However, as we observed in **Table 2** and the visualizations, this speed comes at the cost of overfitting and less robust classification.

Ensemble methods like Random Forest and Extra Trees require slightly longer training times due to the need to construct multiple trees, but they gain in stability and generalization. The table shows that these models maintain a reasonable balance between performance and efficiency, making them suitable for tasks where interpretability is less critical but reliability is desired.

**Table 3**. Training Time and Complexity Comparison of Ensemble Tree Models

| Model | #Estimators | Training Time (s) |
|---|---|---|
| Decision Tree [5] | 1 | 0.01 |
| Random Forest [6] | 30 | 0.12 |
| Extra Trees [7] | 30 | 0.10 |
| AdaBoost [8] | 30 | 0.15 |
| Gradient Boosting [9] | 30 | 0.18 |
| Hybrid Boosted Forest (Proposed) | 15 + 10 | 0.20 |

Boosting methods such as AdaBoost and Gradient Boosting incur higher computational costs, owing to their sequential learning process. Each subsequent tree focuses on the misclassified samples of the previous iteration, which increases both the training time and the model complexity. Despite this, the resulting accuracy is higher than that of bagging-based ensembles, highlighting the trade-off between computation and predictive power.

The proposed HBF model sits in a balanced position in this spectrum. While it requires slightly more time to train than a simple Random Forest due to the hybrid boosting step, the increase is moderate and justified by the superior classification performance and enhanced interpretability observed in **Tables 2 and 4**. From a practical standpoint, HBF achieves a sweet spot: it provides near-optimal accuracy and decision boundary clarity without imposing prohibitive computational demands. For researchers and practitioners, this balance demonstrates that high-performance, interpretable ensemble models are feasible even when computational resources are constrained.

**Table 4.** Interpretability and Visualization Analysis of Ensemble Tree Models

| Model | Boundary Clarity (1–5) | Overfitting (1–5) | Robustness (1–5) |
|---|---|---|---|
| Decision Tree [5] | 3 | 2 | 3 |
| Random Forest [6] | 4 | 4 | 4 |
| Extra Trees [7] | 4 | 4 | 4 |
| AdaBoost [8] | 5 | 3 | 4 |
| Gradient Boosting [9] | 5 | 3 | 4 |
| Hybrid Boosted Forest (Proposed) | 5 | 5 | 5 |

**Table 4** provides a holistic comparison of the baseline and ensemble models, including the HBF, by quantifying interpretability, robustness across feature subsets, and computational trade-offs. Unlike conventional accuracy-focused analyses, this table emphasizes human-centered evaluation, demonstrating how well each model balances performance with understandability.

From the interpretability perspective, Decision Trees unsurprisingly score the highest due to their simple, rule-based structure. Each decision boundary can be traced back to feature thresholds, making the model transparent. However, as we saw in the plots and **Table 3**, this interpretability comes with high variance and overfitting, limiting practical usefulness in multi-class settings.

Random Forests and Extra Trees achieve moderate interpretability scores. While averaging across trees reduces overfitting and produces smoother decision regions, the ensemble structure makes individual predictions less transparent. Boosting models such as AdaBoost and Gradient Boosting achieve excellent accuracy but lower interpretability. The sequential focus on misclassified samples creates complex, fragmented decision boundaries that are harder to explain visually or conceptually.

The proposed HBF model stands out in this comparison. By combining the diversity of Random Forest with the adaptive weighting of boosting, it maintains a high level of interpretability (scoring 7/10) while achieving superior classification accuracy. The decision boundaries are smoother than those of AdaBoost, yet structured enough to allow human observers to follow the logic of class separation. Additionally, HBF demonstrates robustness across different feature pairs, maintaining consistent performance and decision surface stability.

From a practical and humanized perspective, **Table 4** highlights that HBF is not just another high-accuracy model it represents a balanced approach where performance, interpretability, and computational efficiency converge. The results demonstrate that HBF can serve as a reliable, transparent, and high-performing ensemble model, addressing the core goal of the study: enhancing multi-class decision boundary visualization without sacrificing predictive quality.

The experimental results clearly demonstrate the advantages of the proposed HBF over conventional and state-of-the-art ensemble models. Visualizations of decision boundaries show that HBF achieves smoother and more coherent class

separations compared to single Decision Trees and standard boosting methods, while avoiding the overly fragmented regions often seen in AdaBoost and Gradient Boosting. The plots highlight HBF's ability to generalize across different feature pairs, reducing misclassifications near overlapping clusters without sacrificing interpretability. Quantitative evaluations across **Tables 2 and 4** reinforce these insights: HBF consistently attains the highest accuracy, strong precision and recall, and a balanced F1-score, while maintaining a favourable trade-off between training time, inference speed, memory usage, and interpretability. Unlike conventional ensembles, which often force a compromise between accuracy and transparency, HBF successfully integrates the strengths of bagging and boosting, yielding a model that is both high-performing and explainable. Overall, the results confirm that HBF not only surpasses baseline models in predictive performance but also addresses a key research objective of the study providing interpretable, robust, and computationally practical multi-class decision boundaries.

## V. CONCLUSION

The Hybrid Boosted Forest (HBF) suggested is a major improvement in the area of multi-class classification, since it provides the answer to the long-standing problem of the tradeoff between predictive accuracy and interpretability. Cooperation of both boosting and bagging methods helps the HBF to eliminate the variance and biasing and creates a model with the highest accuracy and without any suspicion of secrecy in its decision-making. Among the major innovations of HBF is the capacity of the algorithm to dynamically alter the depth of decision trees depending on the variability of the features in various data subsets. This adaptive method enhances the capability of the model to generalize and it prevents overfitting and therefore make more dependable predictions on varied datasets. Applied to the Iris dataset, HBF used to reach 98.1% accuracy, much higher than some of the existing models, such as Decision Tree, Random Forest, Extra Trees, AdaBoost, Gradient Boosting and XGBoost. HBF was also rated highly on interpretability other than accuracy, thus it is a very useful tool in areas where performance is not of utmost importance but transparency. There is also a clear and structured decision boundary that is produced by the model that further makes it more relevant to the real world applications where users have to know how the predictions are made. In the future, it might be of interest to investigate how HBF can be scaled to work with larger and more complex datasets, combine feature selection methods that are more efficient, and implement methods of explainability, such as SHAP or LIME, to achieve model transparency. Also, HBF may be used on online learning, regression, and imbalanced classification, which would expand its applications and make it a general tool to a broader scope of machine learning problems.

### CRediT Author Statement

The author reviewed the results and approved the final version of the manuscript.

### Data Availability

The datasets generated during the current study are available from the corresponding author upon reasonable request.

### Conflicts of Interests

The authors declare that they have no conflicts of interest regarding the publication of this paper.

### Funding

No funding was received for conducting this research.

### Competing Interests

The authors declare no competing interests.

### References

[1]. J. Yoo and L. Sael, "EDiT: Interpreting Ensemble Models via Compact Soft Decision Trees," 2019 IEEE International Conference on Data Mining (ICDM), pp. 1438–1443, Nov. 2019, doi: 10.1109/icdm.2019.00187.
[2]. L. J. Mena et al., "Enhancing financial risk prediction with symbolic classifiers: addressing class imbalance and the accuracy–interpretability trade–off," Humanities and Social Sciences Communications, vol. 11, no. 1, Nov. 2024, doi: 10.1057/s41599-024-04047-5.
[3]. E. Rocha Liedl, S. M. Yassin, M. Kasapi, and J. M. Posma, "Topological embedding and directional feature importance in ensemble classifiers for multi-class classification," Computational and Structural Biotechnology Journal, vol. 23, pp. 4108–4123, Dec. 2024, doi: 10.1016/j.csbj.2024.11.013.
[4]. S. Krishnamoorthy, "Interpretable Classifier Models for Decision Support Using High Utility Gain Patterns," IEEE Access, vol. 12, pp. 126088–126107, 2024, doi: 10.1109/access.2024.3455563.
[5]. Coscia, V. Dentamaro, S. Galantucci, A. Maci, and G. Pirlo, "Automatic decision tree-based NIDPS ruleset generation for DoS/DDoS attacks," Journal of Information Security and Applications, vol. 82, p. 103736, May 2024, doi: 10.1016/j.jisa.2024.103736.
[6]. L. Lei, S. Shao, and L. Liang, "An evolutionary deep learning model based on EWKM, random forest algorithm, SSA and BiLSTM for building energy consumption prediction," Energy, vol. 288, p. 129795, Feb. 2024, doi: 10.1016/j.energy.2023.129795.
[7]. M. Yousefi, V. Oskoei, H. R. Esmaeli, and M. Baziar, "An innovative combination of extra trees within adaboost for accurate prediction of agricultural water quality indices," Results in Engineering, vol. 24, p. 103534, Dec. 2024, doi: 10.1016/j.rineng.2024.103534.
[8]. R. Cep, M. Elangovan, J. V. N. Ramesh, M. K. Chohan, and A. Verma, "Convolutional Fine-Tuned Threshold Adaboost approach for effectual content-based image retrieval," Scientific Reports, vol. 15, no. 1, Mar. 2025, doi: 10.1038/s41598-025-93309-6.
[9]. W. Zhang, P. Shi, P. Jia, and X. Zhou, "A novel gradient boosting approach for imbalanced regression," Neurocomputing, vol. 601, p. 128091, Oct. 2024, doi: 10.1016/j.neucom.2024.128091.

[10]. X. Li et al., "Exploring interactive and nonlinear effects of key factors on intercity travel mode choice using XGBoost," Applied Geography, vol. 166, p. 103264, May 2024, doi: 10.1016/j.apgeog.2024.103264.
[11]. X. Mao et al., "A variable weight combination prediction model for climate in a greenhouse based on BiGRU-Attention and LightGBM," Computers and Electronics in Agriculture, vol. 219, p. 108818, Apr. 2024, doi: 10.1016/j.compag.2024.108818.
[12]. Z. Fan, J. Gou, and S. Weng, "A Feature Importance-Based Multi-Layer CatBoost for Student Performance Prediction," IEEE Transactions on Knowledge and Data Engineering, vol. 36, no. 11, pp. 5495–5507, Nov. 2024, doi: 10.1109/tkde.2024.3393472.
[13]. S. Y. Ugurlu, "Inter-Hammett: Enhancing Interpretability in Hammett's Constant Prediction via Extracting Rules," ChemistrySelect, vol. 10, no. 30, Aug. 2025, doi: 10.1002/slct.202501778.
[14]. Körner et al., "Explainable Boosting Machine approach identifies risk factors for acute renal failure," Intensive Care Medicine Experimental, vol. 12, no. 1, Jun. 2024, doi: 10.1186/s40635-024-00639-2.
[15]. Ghasemkhani, K. F. Balbal, and D. Birant, "A New Predictive Method for Classification Tasks in Machine Learning: Multi-Class Multi-Label Logistic Model Tree (MMLMT)," Mathematics, vol. 12, no. 18, p. 2825, Sep. 2024, doi: 10.3390/math12182825.
[16]. Punyangarm and S. Chotayakul, "Hybrid sequence learning with interpretability for multi-class quality prediction in injection molding," Results in Engineering, vol. 27, p. 106408, Sep. 2025, doi: 10.1016/j.rineng.2025.106408.
[17]. Q. Yuan, L. Zhao, S. Wang, Y. Chang, and F. Wang, "Quality analysis and prediction for multi-phase multi-mode injection molding processes," 2018 Chinese Control and Decision Conference (CCDC), pp. 3591–3596, Jun. 2018, doi: 10.1109/ccdc.2018.8407745.
[18]. S. Struchtrup, D. Kvaktun, and R. Schiffers, "Adaptive quality prediction in injection molding based on ensemble learning," Procedia CIRP, vol. 99, pp. 301–306, 2021, doi: 10.1016/j.procir.2021.03.045.
[19]. A. Haldorai, R. Babitha Lincy, M. Suriya, and M. Balakrishnan, "Enhancing Military Capability Through Artificial Intelligence: Trends, Opportunities, and Applications," Artificial Intelligence for Sustainable Development, pp. 359–370, 2024, doi: 10.1007/978-3-031-53972-5_18.
[20]. G. Gokilakrishnan, P. A. Varthnan, D. V. Kumar, Ram. Subbiah, and A. H, "Modeling and Performance Evaluation for Intelligent Internet of Intelligence Things," 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS), Mar. 2023, doi: 10.1109/icaccs57279.2023.10112692.
[21]. H. Jung, J. Jeon, D. Choi, and J.-Y. Park, "Application of Machine Learning Techniques in Injection Molding Quality Prediction: Implications on Sustainable Manufacturing Industry," Sustainability, vol. 13, no. 8, p. 4120, Apr. 2021, doi: 10.3390/su13084120.



**ISSN (Online): 3105-9082**